Chapter Six

Writing Examinations

By Adelaide Cherry, Robert Carrigan, and Gary McLean

About the Authors

Dr. Adelaide K. Cherry currently serves as the academic advisor to the commander of the Air Force Institute for Advanced Distributed Learning. She is responsible for overseeing web-based course development, establishing a research and evaluation strategy, and developing a program management service to broker contracts for Air Force customers. Dr. Cherry previously served as the chief of academic affairs for the Air Force Distance Learning Office, chief of educational technology for Air University, and education specialist at the Extension Course Institute. Dr. Cherry received a BS in Secondary Education (Political Science) in 1970, an M.A. in Education (History) in 1974, and a Ph.D. in the Administration of Higher Education in 1977 from the University of Alabama.

Mr. Robert Carrigan is a life long learning advocate with more than 20 years of experience in curriculum development. He is a certified Technical Trainer and earned his teacher certification with the City and Guilds of London, England. He received his Associate of Applied Science (AAS) degree in Aircraft Maintenance Technology in 1989 and his AAS degree in Instructional Technology in 1996 both from the Community College of the Air Force (CCAF). He received his Bachelor of Science degree in Professional Aeronautics from Embry-Riddle Aeronautical University in 1998 and his Master of Science degree in Adult Education from Troy University in 2001. He currently works as an Instructional Systems Specialist at the Air Force Institute for Advanced Distributed Learning (AFIADL) developing text and tests for the Air Force's correspondence education program.

Assessing student performance is getting a lot of attention nowadays. It is so important that DETC has three standards dealing with the subject.

Mr. Gary McLean received his Associate of Applied Science (AAS) degree in Radio Technology in 1986 and his AAS degree in Instructional Technology in 1988 both from the Community College of the Air Force (CCAF). He received his Bachelor of Science degree in Industrial and Vocational Education (IVE) in 1990 and his Master of Science degree in IVE in 1991 both from the University of Southern Mississippi. He is currently working on his dissertation for an Ed.D. in Leadership Education from Spalding University in Louisville, KY. He currently works for the Federal Civil Service as an instructional systems specialist at the Air Force Institute for Advanced Distributed Learning (AFIADL) writing text and tests for the Air Force's correspondence education program.

Introduction

Assessing student performance is getting a lot of attention nowadays. It is so important that DETC has three standards dealing with the subject. The first standard states that examinations and other evaluative techniques must adequately measure achievement of the stated learning objectives; the second states that the institution will provide assessment services that assist students in attaining the intended learning outcomes; and the third requires the institution to demonstrate that students achieve learning outcomes appropriate to its mission and to the rigor and depth of the degrees, diplomas, or certificates offered. In other words, using valid and reliable assessment techniques is serious business for DETC institutions.

This chapter presents information on the purposes of evaluation and why we examine students. It covers the examination development process, some guidelines, a planning matrix, and computer-based testing to help you construct quality examinations.

Purpose of Evaluation

Generally speaking, the effectiveness of an educational institution is determined through an evaluation process that has several purposes—most of which focus on student achievement. A sound evaluation process will help educators, supervisors, and other educational experts:

- Determine student aptitude/ability
- Show student progress toward meeting learning objectives
- Create better learning experiences
- Evaluate the quality of educational programs
- Identify academic strengths of students and programs
- Identify areas of difficulty and make improvements

The effectiveness of an educational institution is determined through an evaluation process that has several purposes—most of which focus on student achievement.

Testing is a vital part of an evaluation program and is one of many tools you may use to evaluate student performance and achievement. Testing is defined as a controlled process for observing a person's performance and rating that performance with the aid of a numerical scale or classification. Testing is a very structured form of student assessment that educators use to determine if a student passes or fails, to monitor student progress, and as an incentive for the student. Testing provides a concrete goal (a grade/score) that can be more motivating than the more elusive goal of "knowledge." It keeps students on track by breaking up the relatively long performance period (training sessions) into discrete performance segments.

Traditionally, student examinations are used to determine placement, promotion, graduation, or retention. A common perception is that tests are used to statistically rank all students according to a sampling of their subject knowledge and to report that ranking to supervisors or anyone else interested in using test information to categorize the students' value. Consequently, tests are highly stressful and anxiety-producing events for most people. Because of this, students should never be measured on the basis of one test alone and test constructors must be careful to develop quality tests.

Examination Development Process

When developing a test, you should consider the elements that impact the reliability and validity of examinations.

When developing a test, you should consider the elements that impact the reliability and validity of examinations. A systematic process that addresses each of these elements is described below, and this process highlights some simple measures you can take to develop quality examinations. The five elements, or steps, in the process include: (1) deciding what to test, (2) determining the value of the test, (3) identifying the required level of student learning, (4) determining the type of test questions to use, and finally, (5) performing a statistical analysis of your test questions. Careful consideration of each step in the process and the use of an evaluation planning matrix will ensure your success.

What to Test

Traditionally, student examinations are focused on learning objectives and educational content that has been identified as important by the community of experts or the learning organization. Test questions must reflect the main objectives within this prescribed curriculum.

For our purposes, let's assume that a training analysis has been completed and the course under development is in the design phase of the instructional development process. Within this phase, course objectives are already identified and you (the test developer) need to develop test questions.

Let's say you want to test the student's mastery of the following objectives through writing test questions:

Given the test development handout, each student will, without instructor assistance:

1. *Recall basic facts and terms associated with the qualitative measures found in the evaluation process.*
2. *Describe the types of tests associated with the levels of learning.*

3. Construct appropriate selection-type questions.

As a test developer, you need to develop questions for each of these objectives. In order to be *valid*, the questions must be congruent with the objectives and the information being referenced. Once you have your objectives, you can begin constructing the evaluation planning matrix by entering each objective in the second column, as shown in Table 1.

Test Value

The next element you need to address is how much emphasis, or value, to place on each objective and the associated test questions. One way to determine this is to look at the amount of time needed for the student to master each lesson objective. In a distance learning environment, the number of test questions should be proportionate to the educational significance of each objective and the material and time allotted for each lesson. For example, let's assume that 20 percent of your examination will be based on recalling basic facts and terms associated with the evaluation process (one-hour lesson), 30 percent on describing the type of tests associated with learning domains (four-hour lesson), and the other 50 percent on writing good questions (five-hour lesson). To ensure an effective evaluation process, all objectives must be met and equal 100 percent.

The left column in the evaluation planning matrix (Table 1) is used to indicate test value.

Table 1

Evaluation Planning Matrix

Percentage/Time	Objectives/Topics	Knowledge	Comprehension	Application	Questions/Points
20%/1 hr	1. Recall basic facts and terms associated with qualitative measures found in the evaluation process.				
30%/4 hrs	2. Describe the types of tests associated with the levels of learning.				
50%/5 hrs	3. Construct appropriate selection-type questions using the guidelines and rules that were presented in class.				
Total = 100%/10 hrs					

Level of Learning

Now that you've determined topical emphasis (value), you are ready to consider the learning level required. You can see that the evaluation planning matrix is a relationship chart that matches the objectives with test value, as well as the level at which you expect students to perform. The chart contains a place for you to indicate which level to use to measure the objective, without emphasizing one objective more than another.

Table 2

Level	Definition	Sample Verbs	Sample Behaviors
Knowledge	Student recalls or recognizes information, ideas, and principles in the approximate form in which it was learned.	Write List Label Name Identify Define	The student will define the 6 levels of Bloom's taxonomy of the cognitive domain.
Comprehension	Student translates, comprehends, or interprets information based on prior learning.	Explain Summarize Paraphrase Describe Illustrate Identify	The student will explain the purpose of Bloom's taxonomy of the cognitive domain.
Level	**Definition**	**Sample Verbs**	**Sample Behaviors**
Application	Student selects, transfers, and uses data and principles to complete a problem or task with a minimum of direction.	Use Compute Solve Demonstrate Apply Construct	The student will write an instructional objective for each level of Bloom's taxonomy.
Analysis	Student distinguishes, classifies, and relates the assumptions, hypotheses, evidence, or structure of a statement or question.	Analyze Categorize Compare Contrast Separate	The student will compare and contrast the cognitive and affective domains.
Synthesis	Student originates, integrates, and combines ideas into a product, plan or proposal that is new to him or her.	Create Design Hypothesize Invent Develop	The student will design a classification scheme for writing educational objectives that combines the cognitive, affective, and psychomotor domains.
Evaluation	Student appraises, assesses, or critiques on a basis of specific standards and criteria.	Judge Recommend Critique Justify	The student will judge the effectiveness of writing objectives using Bloom's taxonomy.

This can be facilitated by using an educational taxonomy, such as the one developed by Benjamin Bloom and others, to categorize domains and levels of learning concerning questions that commonly occur in educational settings. Bloom's taxonomy provides a useful structure to categorize test questions based on objectives. The taxonomy is presented in Table 2 with sample verbs and a sample behavior statement for each level. It is possible to use certain verbs in more than one level of learning.

Whether you use Bloom's taxonomy or another classification system, this chart can help you identify the types of questions (or other types of assessment) that are appropriate. To simplify this portion of the examination development process, let's use the first three levels: knowledge (recall or recognition), comprehension (understanding), and application (demonstration of skill). Yse levels of learning that are relevant to your students and relevant to your instruction. At this stage you can place an "X" in the cells of the evaluation planning matrix to show the levels at which each objective will be measured. Keep in mind that some objectives can be measured at different levels, so questions should be slip accordingly (see Table 3, Objective 3).

Table 3

Percentage/Time	Objectives/Topics	Knowledge	Comprehension	Application	Questions/Points
20%/1 hr	1. Recall basic facts and terms associated with qualitative measures found in the evaluation process.	X			
30%/2 hrs	2. Describe the types of tests associated with the levels of learning.		X		
50%/3 hrs	3. Construct appropriate selection-type questions using the guidelines and rules that were presented in class.		X	X	
Total = 100%/6 hrs					

You can see that Objective 1 (20 percent of the test) is measured entirely at the knowledge level, and Objective 2 (30 percent of the test) at the comprehension level, allowing us to go ahead and enter the total percent of questions we intend to use (Table 4). Since Objective 3 is measured at two different levels, we must decide how to divide 50 percent of the test between comprehension and application levels. You may need to include some questions students may have to explain the difference between selection-type questions (comprehension) and illustrate the ability to develop appropriate examinations (application). Since the emphasis and time given to each level of learning was

different, we may allocate 40 percent to the comprehension part and 10 percent to the application portion of Objective 3.

Table 4

Percentage/Time	Objectives/Topics	Knowledge	Comprehension	Application	Questions/Points
20%/1 hr	1. Recall basic facts and terms associated with qualitative measures found in the evaluation process.	X (20%)			
30%/2 hrs	2. Describe the types of tests associated with the levels of learning.		X (30%)		
50%/3 hrs	3. Construct appropriate selection-type questions using the guidelines and rules that were presented in class.		X (40%)	X (10%)	
Total = 100%/6 hrs					

Type of Test Questions

By now you're probably asking yourself how many and what type of questions should be included in a test. To adequately sample the objectives and the content, you need to consider what type question is best suited to address the required level of learning. You must also consider the amount of training emphasis and time (value) given to each objective. Increasing the number of questions increases the probability that you will be able to adequately assess what the student knows and can do. A good rule of thumb appears in Table 5. This chart, provided by Anthony Nitko, shows that the number of questions and the types of questions used affect the amount of time a student needs to complete the test. Poor readers might need more time.

Table 5

True-False questions	15 to 30 seconds per question
Multiple choice (recall questions that are brief)	30 to 60 seconds
More complex multiple choice questions	60 to 90 seconds
Multiple choice problems with calculations	2 to 5 minutes
Short answer (one word)	30 to 60 seconds
Short answer (longer than one word)	1 to 4 minutes
Matching (5 premises, 6 responses)	2 to 4 minutes
Short essays	15 to 20 minutes
Data analyses/graphing	15 to 25 minutes
Drawing models/labeling	20 to 30 minutes
Extended essays	35 to 50 minutes

Use these estimates to decide what type and how many questions you use. You can see that more true-false questions can be answered during a given period of time than multiple choice or short answer questions. Your question types, however, must be based on the level of learning at which you are evaluating students. True-false and matching questions can assess the knowledge component, and multiple choice questions can assess comprehension.

Test questions can be written in any domain of instruction but will fall into one of three main categories: selection, supply, and demonstration/ performance. They will also include five main question types. Each type of test question has its own set of strengths and weaknesses, as shown in Table 6 below. If you are using multiple choice questions, you may want to consult the guidelines for writing effective examinations that appeared in the 1993 *DETC Course Development Handbook*.

Table 6

Test Type	Question Type	Advantages	Disadvantages	Best Level of Learning
Selection	Multiple Choice	➢ Can measure most levels of student ability. ➢ Enables wide sampling of subject content. ➢ Quick and easy to score. ➢ Enables objective scoring. ➢ Can be analyzed for effectiveness.	➢ Difficult to construct good items. ➢ Tendency to measure simple recall.	➢ Knowledge ➢ Comprehension ➢ Application ➢ Analysis*
	Matching	➢ Relatively easy to construct. ➢ Conserves examinees' reading time. ➢ Enables efficient and objective scoring.	➢ Generally unsuitable for testing higher-order abilities. ➢ Tendency to measure simple recall.	➢ Knowledge ➢ Comprehension
	True/False	➢ Efficient for testing large sample of information. ➢ Enables efficient and objective scoring.	➢ Permits high guess factor. ➢ Difficult to construct effective items.	➢ Knowledge ➢ Comprehension
Supply	Completion/ Short Answer	➢ Minimizes guessing. ➢ Enables coverage of fairly wide content. ➢ Relatively easy to construct.	➢ Measures limited range of abilities. ➢ Cannot be machine-scored. ➢ Scoring is highly dependent on judgment.	➢ Knowledge ➢ Comprehension ➢ Application ➢ Analysis*
	Essay	➢ Can be quickly and easily constructed. ➢ Eliminates guessing. ➢ Can test higher order of thinking.	➢ Limits amount of content sampled. ➢ Time-consuming to score. ➢ Results in low scoring reliability.	➢ Analysis ➢ Synthesis ➢ Evaluation

* Can be reached with considerable experience and practice of writing this type of question.

Very often, the "logical" choice of any test format must be tailored to accommodate practical considerations such as class size, facilities and resources, time constraints, and other environmental issues.

Very often, the "logical" choice of any test format must be tailored to accommodate practical considerations such as class size, facilities and resources, time constraints, and other environmental issues.

Referring to the evaluation planning matrix, you can see that questions pertaining to Objective 3 may be time-consuming for students. Even though this is a paper-and-pencil test, Objective 3 is basically a performance task that will require measurement at the application level of learning. Therefore, you need a weighted question system, which usually entails assigning lower points to lower levels of learning. For example, award the lowest number of points per question (one point per question) for the easiest questions (in this case the true-false and matching) and increase point value as learning difficulty increases. Weights will help you determine the number of true-false, matching, multiple choice, and other type questions to include on the test.

To finish your evaluation matrix, put the number of proposed test items in the appropriate cell. By viewing the completed evaluation matrix below, you can see the total number of questions at the far right of each column. You can also see that your test needs 10 knowledge questions (true-false and matching type questions are good for evaluating the lowest level of learning "knowledge") worth 1 point each. Since 35 comprehension questions are needed, you must select a type of question that will challenge the student a bit more. Multiple choice questions can meet this challenge if developed properly. Since the questions are more difficult, more points should be given, and in this case two points are given for each comprehension type question (see Objectives 2 & 3). Since five application level questions needed, use scenario-based or essay-style questions. Since application level is more difficult than comprehension level, more points should be given to the student for answering this type of question correctly. As you can see from the table, four points are given to each question.

The amount of questions and points allocated to each question in the right-hand column gives us a total of 100 points in a 50 question test.

Allow enough time for the slowest students to complete your test, and make sure the test can be taken in a single session.

When estimating the time needed for this test, students will probably require three to five minutes for the 10 true-false questions (15–30 seconds each), about 30 minutes for the 35 comprehension questions (60–90 seconds each), and 30–40 minutes (a rough estimate) to read the material and answer the five short answer questions measuring the application level. The total time needed would be from one to one and a half hours. Allow enough time for the slowest students to complete your test, and make sure the test can be taken in a single session.

Table 7

Percentage/Time	Objectives/Topics	Knowledge	Comprehension	Application	Questions/Points
20%/1 hr	1. Recall basic facts and terms associated with qualitative measures found in the evaluation process.	X (20%) (10 Q)			10 Q/10 pts (1 pt each)
30%/2 hrs	2. Describe the types of tests associated with the levels of learning.		X (30%) (15 Q)		15 Q/30 pts (2 pts each)
50%/3 hrs	3. Construct appropriate selection-type questions using the guidelines and rules that were presented in class.		X (40%) (20 Q)	X (10%) (5 Q)	20 Q/40 pts (2 pts each) 5 Q/20 pts (4 pts each) (Total = 25 Q/60 pts)
Total = 100%/6 hrs		10	35	5	50 Q/100 pts

Test Validation (statistical analysis)

The final step in the process is to validate the test by conducting a statistical analysis. This determines whether your test is performing as needed. The information gained from a statistical analysis will provide a means for improving subsequent exams. Fortunately, data can be calculated during the analysis of selection type exams that provide hints on how to improve an exam.

Your overall aim in an item analysis of selection type questions is to increase the reliability and validity of the test.

Your overall aim in an item analysis of selection type questions is to increase the reliability and validity of the test. Analyzing a set of questions is characterized in a test statistics report or other statistical collection process. Computerized statistical packages can automate this process and generate a report that gives you an overall view of the test results. These results measure the central tendency, variability, and shape of the test score distribution within a test taking population. The central concepts are item difficulty level, item discrimination, distracter analyses, item-test intercorrelations, and a variety of related statistics that show the overall test difficulty.

Classical test analysis also includes a measure for the reliability of scores. Understanding the significance of the statistical data and knowing how to improve your test is important.

The statistical analysis "tests the test." The test statistics report analyzes each test item and overall test reliability. A proper item analysis points out items that are miskeyed, ambiguous, overly easy or difficult, and nondiscriminating. These items detract from overall test reliability. Item analysis also shows the kinds of errors students make on test items.

Two factors determine overall test effectiveness: test validity and test reliability.

Two factors determine overall test effectiveness: test validity and test reliability. Test validity is the appropriateness of the test for the subject area and students being tested. A computer cannot measure validity. It is up to the educator to use statistical measures to design valid test items that best measure the intended subject area. By definition, valid tests are reliable. However, a reliable test is not necessarily valid. For example, a math test comprised entirely of word problems may be measuring as much verbal skills as math ability. The relationship between validity and reliability has been described in many textbooks and can be better explained through a picture.

A **B** **C**

Imagine you have three tests, each test comprised of 100 questions. How would you know if the tests are reliable and valid? A favorite metaphor for the relationship between reliability and validity is the target concept. Think of the center of the target (bull's-eye) as the objective you are trying to measure. Imagine you place a mark on a graded target for each person you measure. If the objective is clear and measured perfectly, then more marks are hitting the center of the target (getting the question right). If each person does not meet the objective, then the person is missing the center.

From the example above, you can see how each test performed differently. The pattern on target (A) shows that this test is not very reliable (it doesn't hit the same place on the target consistently). In fact, this test doesn't even come close to measuring the learning objective. Student answers are all over the target. The pattern on target (B) shows the marks are focused in a small area on the target. We could say this test is reliable because it consistently hits the same place on the target, but it still isn't valid because it doesn't hit the bull's-eye. Only target (C) displays a pattern that shows high reliability and validity.

Test reliability measures the accuracy, stability, and consistency of the test scores. Reliability is affected by the characteristics of the students, the characteristics of the test, and the conditions affecting test administration and scoring. Reliability is usually expressed as some sort of correlation coefficient with a values range from .00 (low reliability) to 1.00 (perfect reliability).

Test reliability measures the accuracy, stability, and consistency of the test scores. Reliability is affected by the characteristics of the students, the characteristics of the test, and the conditions affecting test administration and scoring.

To determine effectiveness of course material, most educational institutes rely on content validity. Content validity describes how well the sample of test items represents the designed test content. This type of validity is achieved through systematically defining, in precise terms, the learning objectives, the specific content, and how the content will be sampled.

Always remember that validity is a process and not a product. Developing a valid and reliable test is very important to you and the student; however, if the test is not administered properly, then all your development and meticulous planning could go to waste. There are many ways to administer a test, but the one becoming more and more common is the computer-based test (CBT).

Computer-Based Tests and Examinations

Traditionally, examinations were given via paper and pencil in a group setting at a particular time. In the past several years, however, more and more organizations have shifted to computerized tests. Using computers for student proficiency testing has become increasingly common. In numerous settings around the United States, computer-based tests are being implemented and important decisions are being made about examinees on the basis of these tests. For example, the Educational Testing Service's (ETS) Graduate Record Examination (GRE) is now offered in a computer-based format, and a CBT version of the Scholastic Achievement Test (SAT) will soon become operational. In addition, computer-based licensure and certification tests have been used for a number of years in many professional fields.

Computer-based testing enables educators to develop types of tests and items that are simply not possible with paper tests.

Applying computer technology to testing gives you the ability to open up the power of testing. Computer-based testing enables educators to develop types of tests and items that are simply not possible with paper tests. Decisions to use CBTs to assess student outcomes should be based on a consideration of all of the relevant issues, advantages, and disadvantages of CBTs.

Advantages of CBTs

Using computer-based testing is attractive for many reasons. First, CBTs can be more responsive to the needs of both the test developer and the student. In contexts where "on-demand" testing of examinees is needed, the use of

computerized tests does not require scheduled administration times. Also, during testing, examinees can receive feedback regarding their progress. Such feedback might take the form of item feedback regarding answer correctness, a running total score, or simply information on how many items are left or how much time is remaining in the testing period.

Moreover, CBT scoring is immediate, which facilitates decisions concerning the basis of an examinee's test score. An additional advantage is the ability to test physically disabled examinees for which administration of conventional tests is infeasible.

A computer-based version of the test item is likely a purer measure of the students' understanding of the electrical concept because it is less confounded with other skills such as reading proficiency.

A second, and potentially more exciting, advantage of CBTs is the capability to present items in new and perhaps more realistic ways. For example, on a conventional career development course, a complex text description or a series of electrical diagrams may be required to present an item concerning a particular property of resistance. On a CBT, the motion in question might be more simply and clearly depicted using a brief video clip or an animated sequence demonstrating the electrical process. A computer-based version of the test item is likely a purer measure of the students' understanding of the electrical concept because it is less confounded with other skills such as reading proficiency. This implies that inferences made regarding student knowledge and skills are potentially more valid under a computer-based format.

Being able to introduce audio and video into test items is a very attractive feature of CBTs. Innovative tests may also create a simulated environment in which students are asked to demonstrate particular proficiencies. For example, a virtual chemistry lab could be set up to allow students to show that they know how to "make" a particular compound. Such computer-based performance assessments should be appealing to many academic programs. The computer provides possibilities for new methods to assess student outcomes, and a great deal of innovation should emerge in the coming years.

A more advanced type of CBT is the computerized adaptive test (CAT). These tests use items programmed to respond differentially to student responses. The adaptive program's mechanisms trace students' pathways through problem and diagnostic items. Items delivered by computer can be randomized for each administration, which can reduce cheating.

In a CAT, the computer adjusts the characteristics of the administered items to match the proficiency level of each examinee. That is, examinees higher in proficiency will receive a more difficult set of items than those lower in proficiency. Because each examinee receives a unique set of items that is matched to his or her proficiency level, testing is more efficient and can be

completed much more quickly. A CAT typically requires about half as many items as a conventional test to attain the same precision of measurement (with commensurate savings in testing time), which has made it very attractive to large-scale testing programs. For instance, the GRE uses a CAT format, as will the SAT when it is offered in a computer-based format.

Disadvantages of CBTs

The disadvantages of computer-based testing are all related to the required resources.

The disadvantages of computer-based testing are all related to the required resources. Basically, these resources include a sufficient number of computers, a room to install them, appropriate software, and adequate technical expertise. The required number of computers comes from two factors: the total number of students and the time required to complete the testing. Thus, lengthening the time frame can compensate for a smaller number of computers.

It might be challenging to find a large enough room to contain the computers. Often, a computer laboratory is developed that can be used for other purposes (instruction, student projects, e-mail) when it is not needed for testing. Such a multi-use room is an efficient way for departments to use limited space.

Obtaining good testing software is a common obstacle when adopting computer-based testing. You may either develop your own software or purchase software from a commercial software vendor. If you develop your own, you can get exactly what you want, but you have to locate and pay a programmer to set it up for you. Commercially available software is often less expensive, but you may have difficulty finding software that matches your specific needs. A hybrid option is to hire a software company to modify its existing software to meet your particular needs. As you might expect, however, this is more expensive than purchasing off-the-shelf software.

A final key element to adopting computer-based testing is having access to an individual with substantial technological expertise.

A final key element to adopting computer-based testing is having access to an individual with substantial technological expertise. Such a person will be very useful in planning the testing room, purchasing the most appropriate computers, making decisions involving software acquisition, and maintaining computer hardware and software once CBTs have been implemented.

Summary

Evaluation is an important method of gauging how much students are actually learning. It is important to periodically evaluate students to determine if and how much they are learning. Educators can use tests to determine whether their instruction was effective, and to discover which students need additional help. Students can use test information to find out if their study habits are paying off. Because a student's intellectual, social, and physical capacity is continuously developing, student evaluation is an ongoing task.

Tests are constructed to elicit evidence of student learning in relation to the instructional objectives.

Tests are constructed to elicit evidence of student learning in relation to the instructional objectives. As test designers, you need to keep the instructional objectives in mind when developing valid examinations. Valid examinations can be obtained if you concern yourself with issues such as what needs to be tested, test value, level of student learning, type of test question to use, and finally, a statistical analysis process to validate the test. Developing an evaluation matrix helps to plan tests that correspond to instructional objectives. Types of test questions include multiple choice, true-false, completion, matching, short essay, and so forth. Each type of test item has optional uses, advantages, and disadvantages. Properly developed tests will increase the reliability and validity of the course.

Without a valid test to measure a learning outcome, you'll never know if you are wasting time in the evaluation process.

As you develop tests, consider any advancement in technology that can enhance your students' learning ability. Technology, like computer-based testing, can be more responsive to the needs of both the test developer and the student. A CBT program enables an "on-demand" testing environment and reduces the need of many administrative procedures. Also, during testing, examinees can receive feedback and challenge questions regarding their test administration.

As you can see, test development is an art requiring skill, patience, and understanding of educational goals and student needs. Without a valid test to measure a learning outcome, you'll never know if you are wasting time in the evaluation process.

Bibliography

Assessment of Student Performance, Volume 1. (ERIC Document Reproduction Service No. Ed 324 204).

Bloom, Benjamin, S. et al. *Taxonomy of Educational Objectives,* Book 1, *Cognitive Domain* (26th ed). New York: Longman Inc, 1983.

Cherry, Adelaide K. *Writing Examinations*, Course Development Handbook. DETC publications, 1993.

Nitko, Anthony J. *Educational Assessment of Students* (2nd ed.). New York: Prentice Hall, 1996.